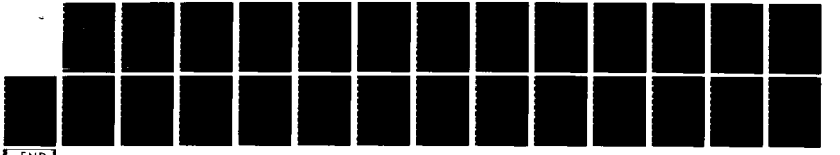
AD-A137 942 ON SEGMENTATION OF DIGITAL IMAGES USING SPATIAL AND CONTEXTUAL INFORMATIO..(U) ILLINOIS UNIV AT CHICAGO CIRCLE DEPT OF QUANTITATIVE METHODS S L SCLOVE 1/1

UNCLASSIFIED 06 DEC 83 UIC/DQM/A83-2 ARO-19085.6-MA F/G 12/1 NL

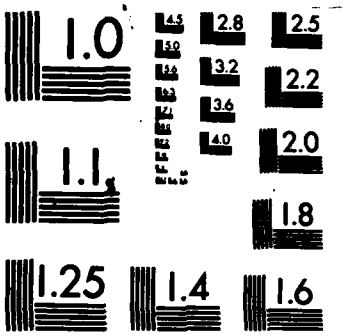


END

FILMED

3-84

DTIC



ON SEGMENTATION OF DIGITAL IMAGES USING SPATIAL AND CONTEXTUAL INFORMATION VIA A TWO-DIMENSIONAL MARKOV MODEL

by

STANLEY L. SCLOVE

A Presentation to the
Workshop on Applied Pattern Recognition and Image Analysis,
sponsored by the Army Research Office
held at Brown University, Providence, Rhode Island, April 14-16, 1983

To appear in
Proceedings of the Conference on Unsupervised Image Analysis,
David B. Cooper, Robt. Launer, and Donald McClure (eds.),
Academic Press, 1984

TECHNICAL REPORT NO. UIC/DQM/A83-2
December 6, 1983

PREPARED FOR THE
ARMY RESEARCH OFFICE
UNDER
CONTRACT DAAG29-82-K-0155

Statistical Models and Methods for
Cluster Analysis and Image Segmentation

Principal Investigator: Stanley L. Sclove

QUANTITATIVE METHODS DEPARTMENT
COLLEGE OF BUSINESS ADMINISTRATION
UNIVERSITY OF ILLINOIS AT CHICAGO
BOX 4348, CHICAGO, IL 60680

12/5/83

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

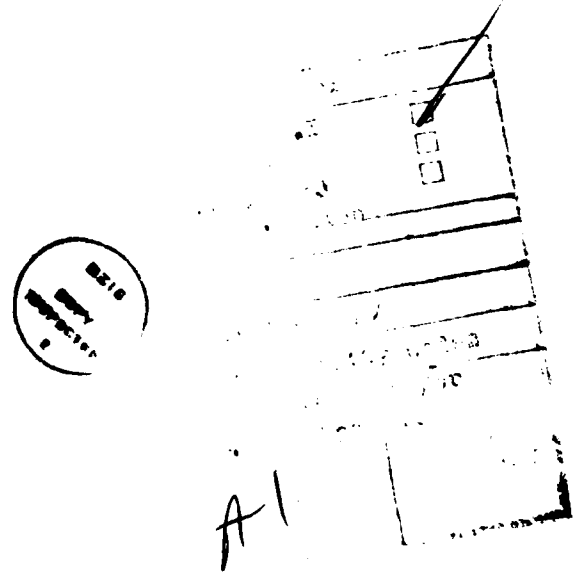
ON SEGMENTATION OF DIGITAL IMAGES USING SPATIAL AND CONTEXTUAL INFORMATION VIA A TWO-DIMENSIONAL MARKOV MODEL

STANLEY L. SCLOVE

Department of Quantitative Methods

College of Business Administration
University of Illinois at Chicago

CONTENTS



ON SEGMENTATION OF DIGITAL IMAGES USING SPATIAL AND CONTEXTUAL INFORMATION VIA A TWO-DIMENSIONAL MARKOV MODEL

STANLEY L. SCLOVE
Department of Quantitative Methods
College of Business Administration
University of Illinois at Chicago

ABSTRACT

The problem of partitioning a digital image into segments is considered. First the procedure is illustrated for the analogous one-dimensional problem, namely, segmentation of time series. Then similar ideas are applied to the segmentation of digital images.

The segments are considered as falling into classes. A probability distribution is associated with each class of segment. Parametric families of distributions are considered, a set of parameter values being associated with each class. With each observation is associated an unobservable label, indicating from which class the observation arose. The label process is modeled as a Markov process. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood (a relaxation method) to the resulting likelihood function. In this paper special attention is given to the situation in which the observations are conditionally independent, given the labels. Numerical examples are given. Choice of the number of classes, using statistical model-selection criteria, is illustrated.

Key words and phrases: time-series segmentation, digital image segmentation, relaxation method, Viterbi algorithm, model-selection criteria, information criteria, Akaike's information criterion, Schwarz' criterion

ON SEGMENTATION OF DIGITAL IMAGES USING SPATIAL AND CONTEXTUAL INFORMATION VIA A TWO-DIMENSIONAL MARKOV MODEL

by

STANLEY L. SCLOVE

Department of Quantitative Methods

College of Business Administration
University of Illinois at Chicago

1. Introduction

Problems of segmenting time series and digital images are considered. In both cases the observations fall into classes, and this assignment of observations to classes, or "labeling," is unknown. Thus each point gives rise to a pair, the observation itself, together with its unknown label. In the context of this model, segmentation is estimation of the labels. In time series the points are time points; in digital images, the points are points of the image (picture elements, or pixels). An image is two-dimensional, while time is one-dimensional, so time series are treated first.

2. Segmentation of Time Series

The problem of segmentation considered here is: Given a time series

$$\{x_t,\ t=1,2,\ldots,n\},$$

partition the set of values of t into sub-series (segments, regimes) in some meaningful way. The segments are assumed to fall into several classes. The classes may be associated with different generating mechanisms. In cyclic processes the classes are phases of the cycle.

Examples. (i) Segment a received signal into background, target, background again, another target, etc. (ii) Segment an EEG of a sleeping person into periods of deep sleep and restless or fitful sleep (two classes of segment). (iii) Segment an ECG into rhythmic and arhythmic periods (two classes of segment). (iv) Segment an economic time series into periods of recession, recovery, and expansion. Here there are three classes of segment.

In some applications the observation X is a vector of several measurements. E.g., for blood pressure, X is a vector of the two measurements, systolic and diastolic. The discussion of segmentation of digital images, later in the paper, will be in terms of vector measurements. Time series will be discussed in terms of scalar (single) measurements, although the ideas and methods readily generalize to multiple time series.

2.1. The Model

One can imagine a series which is usually relatively smooth but occasionally rather jumpy as being composed of sub-series which are first-order autoregressive, the autocorrelation coefficient being positive for the smooth segments and negative for the jumpy ones. One might try fitting such data with a segmentation of two classes, one corresponding to a positive autocorrelation, the other, to a negative autocorrelation.

The mechanism generating the process changes from time to time, and these changes manifest themselves at some unknown time points (epochs, or change-points)

$$t_1, t_2, \ldots, t_{m-1};$$

i.e., the number of segments is m. The integer m and the epochs are unknown. Generally there will be fewer than m generating mechanisms. The number of mechanisms (classes) will be denoted by k; it will be assumed that k is at most m. In some situations, k is specified; in others, it is not. With the c-th class is associated a stochastic process, P_c, say. E.g., above we

spoke of a situation with k=2 classes, where, for c=1,2, the process P_c is first-order autoregressive with coefficient, say, ϕ_c.

Now with the t-th observation (t=1,2,...,n) associate the label γ_t, which is equal to c if and only if x_t arose from class c, c = 1,2,...,k. Each time-point t gives rise to a pair

$$(x_t, \gamma_t),$$

where x_t is observable and γ_t is not. The process $\{x_t\}$ is the observed time series; $\{\gamma_t\}$ will be called the "label process."

Define a segmentation, then, as a partition of the time index set $\{t: t = 1,2,...,n\}$ into subsets

$S_1 = \{1,2,...,t_1\}$, $S_2 = \{t_1+1,...,t_2\}$,..., $S_m = \{t_{m-1}+1,...,n\}$,

where the t's are subscripted in ascending order. Each subset S_g, g = 1,2,...,m, is a segment. The integer m is not specified. In the context of this model, to segment the series is merely to estimate the γ's.

The idea underlying the development here is that of transitions between classes. The labels γ_t will be treated as random variables in a stochastic process with transition probabilities

$$Pr(\gamma_t=d|\gamma_{t-1}=c) = p_{cd},$$

taken as stationary, i.e., independent of t. The k-by-k matrix of transition probabilities will be denoted by $\underline{P}$, i.e.,

$$\underline{P} = [p_{cd}].$$

If a process is strictly cyclic, like intake, compression, combustion, intake, etc., for a combustion engine, or recession to recovery to expansion to recession, etc., in the business cycle, then this condition can be imposed by using a transition probability matrix with zeros in the appropriate places. We shall consider such a matrix in Section 2.3.2.

Segmentation will involve the simultaneous estimation of several sets of parameters, the distributional parameters of the within-class stochastic processes, the transition probabilities, and the labels.

A joint probability density function (p.d.f.) for $\{(X_t,\gamma_t),\ t=1,2,\ldots,n\}$ is obtained by successively conditioning each variable on all the preceding ones. The working assumptions behind the method of this paper are as follows.

(A.1) The labels are a first-order stationary Markov chain, independent of the observations; i.e., the probability of being in state d at time t+1 given state c at time t, is p_{cd}, which does not involve t or the values of the observations.

(A.2) The distribution of the random variable X_t depends only on its own label and previous X's, not previous labels.

Further details and a mathematical formulation corresponding to these assumptions are given in Sclove (1983a).

In regard to (A.2), in the simplest case the X's are (conditionally) independent, given the labels. That is, the disribution of X_t depends only on its label, and not previous X's. In the examples in the present paper this assumption is made. In this case the p.d.f.'s $f(x|\gamma_t=c)$, $c = 1,2,..,k$, are called <u>class-conditional</u> densities. In the parametric case they take the form

$$f(x_t|\gamma_t=c) = g(x_t;\beta_c), \qquad (1)$$

where β is a parameter indexing a family of p.d.f.'s of form given by the function g. E.g., in the case of Gaussian class-conditional distributions, β_c consists of the mean and variance for the c-th class.

This model, with transition probabilities, has certain advantages over a model formulated in terms of the epochs. The epochs behave as

constrained discrete parameters, and, even if the suitably transformed generalized likelihood ratio were asymptotically chi-square, the number of degrees of freedom would not be clear. On the other hand, the transition probabilities vary in an interval and it is clear that they constitute a set of $k(k-1)$ free parameters.

2.2. An algorithm

2.2.1. Development of the algorithm

It follows from the assumptions that the likelihood L, i.e., the joint p.d.f, considered as a function of the parameters, can be written in the form

$$L = A(\{p_{cd}\}, \{\gamma_t\})B(\{\gamma_t\}, \{\beta_c\}). \qquad (2)$$

Hence, for fixed values of the γ's and β's, L is maximized with respect to the p's by maximizing the factor A. Now let n_{cd} be the number of transitions from class c to d. (These n's are functions of the labels.) The factor A is merely the point multinomial probability function, the parameters being the n's and p's. It follows that the maximum likelihood estimates of the p's, for fixed values of the other parameters, are given by

$$p_{cd} = n_{cd}/n_c , \qquad (3)$$

where

$$n_c = n_{c1} + n_{c2} + \ldots + n_{ck} .$$

Further, given the p's and γ's, the estimates of the distributional parameters--the β's--are easy to obtain. This suggests the following algorithm.

Step 0. Set the β's at initial trial values, suggested, e.g., by previous knowledge of the phenomenon under study. Set the p's at initial trial values, e.g., $1/k$. Set $f(\gamma_1)$ at initial trial values, e.g., $f(\gamma_1) = 1/k$, for $\gamma_1 = 1,2,\ldots,k$.

Step 1. Estimate γ_1 by maximizing $f(\gamma_1)f(x_1|\gamma_1)$.

Step 2. For t=2,3,...,n, estimate γ_t by maximizing

$$p_{\gamma_{t-1}\gamma_t} f(x_t|\gamma_t, x_{t-1}, \ldots, x_1),$$

as, under (A.1) and (A.2), the likelihood can be expressed as a product of such factors.

Step 3. Now, having labeled the observations, estimate the distributional parameters, and estimate the transition probabilities according to (3).

Step 4. If no observation has changed labels from the previous iteration, stop. Otherwise, repeat the procedure from Step 1.

This algorithm is a "relaxation method," in the sense of Southwell (1940, 1946). That is, one maximizes successively on each variable, holding the others fixed.

Having cast the problem in statistical terms and obtained a likelihood, one can score successive iterations according to the likelihood values. This solves the problem of which iteration's results to use, when there is not always improvement from iteration to iteration. (See, e.g., Richards, Landgrebe and Swain (1980) for a discusssion of this problem with relaxation methods for pixel labeling.)

To maximize by considering the factors mentioned in Step 2 is to use the Viterbi algorithm (Forney 1973). (The Viterbi algorithm is a recursive optimal solution to the problem of estimating the state sequence of a discrete-time finite state Markov process.)

Step 2 is Bayesian classification of x_t. Suppose the (t-1)-st observation was tentatively classified into class c. Then the prior probability that the t-th observation belongs to class d is p_{cd}, d = 1,2,...,k. Hence all the techniques for classification in particular models are available (e.g., use of linear discriminant functions when the observations are multivariate normal with common covariance matrix).

2.2.2. The first iteration

When the k class-conditional processes consist of independent, identically distributed normally distributed random variables with common variance, one can start by choosing initial means and labelling the observations by a minimum-distance clustering procedure. [This is one iteration of ISODATA (Ball and Hall, 1967). One could iterate further at this stage.] From this clustering initial estimates of transition probabilities and the variance are obtained.

2.2.3. Restrictions on the transitions

As mentioned above, one might wish to place restrictions on the transitions, e.g., to allow transitions only to adjace states. (E.g., "recovery" is adjacent to "recession", "expansion" is acent to "recovery," but "expansion" is not adjacent to "recession.") The m e loes permit restrictions on the transitions. The maximization is conducted, subject to the condition that the corresponding transition probabilities are zero. This is easily implemented in the algorithm. Once one sets a given transition probability at zero, the algorithm will fit no such transitions, and the corresponding transition probability will remain set at zero at every iteration.

2.3. An example

Here, in the context of a specific numerical example, the problem of fitting the model for a fixed k and the problem of choice of k will be illustrated. (The additional problem of prediction is considered in Sclove 1983a.)

Quarterly gross national product (GNP) in current (non-constant) dollars for the twenty years 1947 to 1966 was considered. (This makes a good size dataset for expository purposes here.) Parameters were estimated from the

first 19 years, the last four observations (1966) being saved to test the accuracy of predictions. (See Sclove 1983a.) The data and first difference are given in Table 1. The raw series is nonstationary, so the first differences (increases in quarterly GNP) were analyzed. (Study of more recent data suggested use of first or second differences of logs; see Sclove (1983b).) The notation is

$$x_t = GNP_{t+1} - GNP_t \ , \ t = 1,2,\ldots,79;$$

e.g., GNP_1 is the GNP at the end of the quarter 1947-1, GNP_2 is that at the end of 1947-2, and $x_1 = GNP_2 - GNP_1$ is the increase in GNP during the second quarter of 1947. (A negative value of an x indicates a decrease in GNP for the corresponding quarter.) A Gaussian model was used.

2.3.1. Fitting the model

In this section we discuss the fitting of a model with k=3 classes, discussion of the choice of k being deferred to the next section. The three classes may be considered as corresponding to recession, recovery, and expansion, although some may prefer to think of the segments labeled as recovery as level periods corresponding to peaks and troughs. The approximate maximum likelihood solution found by the iterative procedure was (units are billions of current (non-constant) dollars) -1.3, 6.2, and 12.3 for the means, 2.28 for the standard deviation, and

.625	.250	.125
.156	.625	.219
.039	.269	.692

for the transition probabilities. The estimated labels are given below; labels (r = recession, e = expansion) resulting from fitting k=2 classes (see Section 2.3.2) are also given.

t:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
label, k=3:	2	2	3	2	2	2	1	1	1	1	1	3	3	3	3	3	2	2	2	2	1	2	3
label, k=2:	r	r	e	e	e	e	r	r	r	r	r	e	e	e	e	e	e	e	e	r	r	e	e

24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
2	2	1	1	1	1	2	2	3	2	2	2	2	2	2	2	2	2	2	1	1	2	3	3	3	3	1
e	r	r	r	r	r	r	e	e	e	e	e	r	r	r	e	e	r	e	r	r	r	e	e	e	e	r

51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
2	3	2	1	1	1	3	3	3	3	3	2	2	2	2	3	3	3	3	3	2	3	3	3	3
e	e	r	r	r	r	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e

The process was in state 1 for 21% of the time, in state 2 for 44% of the time, and in state 3 for 35% of the time. The labeling corresponds well to the conventional wisdom of economists concerning recession during these years; see Sclove (1983a).

An interesting feature of the model and the algorithm is that, as the iterations proceed, some isolated labels change to conform to their neighbors. This should be the case when p_{cc} is large relative to p_{cd}, for d different from c.

2.3.2. Choice of number of classes

Various values of k were tried, the results being scored by means of Akaike's information criterion (AIC). (See, e.g., Akaike 1981.) To estimate k one uses the minimum AIC estimate, where

$$AIC(k) = -2\log_e[\max L(k)] + 2m(k).$$

Here L(k) is the likelihood when k classes are used, max denotes its maximum over the parameters, and m(k) is the number of independent parameters when k classes are used. The statistic AIC(k) is a natural estimate of the "cross-entropy" between f and g(k), where f is the (unknown) true density and g(k) is the density corresponding to the model with k

classes. (See, e.g., Parzen (1982) for a discussion of the cross-entropy, H(f;g).) According to AIC, inclusion of an additional parameter is appropriate if $\log_e$[max L] increases by one unit or more, i.e., if max L increases by a factor of e or more. It may be generally preferable to consider the Schwarz (1978) criterion

$$-2\log_e[\max L(k)] + (\log_e n)m(k).$$

It favors models with fewer parameters. However, in this particular example, as will be seend, even AIC favors the model with thn minimum number of parameters.

The model was fit with several values of k and unrestricted transition probabilities. Also, since it seems reasonable to restrict the transitions to those between adjacent states, such models were evaluated as well. In the case of k=3, where the states might be considered as recession, recovery, and expansion, this means setting equal to zero the transition probabilities coresponding to the transitions, recession-to-expansion and expansion-to-recession. The results are given in Table 2. The best segmentation model, as indicated by minimum AIC, is that with only two classes.

A model with only two classes enjoys advantages relating to its relative simplicity.

The results for k=2 classes (which might be called recession, expansion) were 0.43 and 10.09 for the means, 3.306 for the standard deviation, and

.667	.333
.170	.830

for the transition probabilities. The process was in state 1 for 37% of the time and state 2 the other 63% of the time. The labels were given above.

The algorithm has been used with good success to segment bovine temperature time series into menstrual and non-menstrual periods and to segment a time series of influenza deaths into epidemic and non-epidemic periods.

3. A Model and Method for Segmentation of Digital Images

A digital (i.e., numerical) image may be considered as a rectangular array of picture elements (pixels). These will be indexed by (i,j). At each pixel the same p features are observed. We denote the features by

$$X_1, X_2, \ldots, X_p.$$

The vector of features is

$$\underline{X} = (X_1, X_2, \ldots, X_p).$$

The digital image is

$$\{\underline{X}_{ij},\ i=1,2,\ldots,I,\ j=1,2,\ldots,J\},$$

where

$$\underline{X}_{ij} = (X_{1ij}, X_{2ij}, \ldots, X_{pij})$$

is the vector of numerical values of the p features at pixel (i,j).

Examples. (i) In television, we have p = 3 colors,
x_{1ij} = red level at pixel (i,j),
x_{2ij} = green level at pixel (i,j),
and
x_{3ij} = blue level at pixel (i,j).

(ii) In LANDSAT data, p = 4 spectral channels, one in the green/ yellow visible range, the second in the red visible range, and the other two in the near infrared range.

An object is a set of contiguous pixels which may be assumed to be members of a common class. One task of image processing is segmentation, grouping of pixels with a view toward identifying objects.

In this context the conceptual model is that the image is a set of

pixels, and, also, the image consists of several segments. Each pixel belongs to one and only one segment. The segments fall into several classes. For example, in a picture of a house the classes might be brick, sky, grass, shadow and brush. Note that there might be several separate areas of, say, grass. Each of these areas is a segment, but they all belong to the class called "grass."

The statistical model accompanying this conceptual model is as follows:

--with each class of segment is associated a probability distribution for the feature vector $\underline{X}$;
--with each pixel is associated a label which, were it known to us, would tell us which class of segment the pixel belongs to.

Each pixel thus gives rise to a pair $(\underline{X},\gamma)$, where $\underline{X}$ is observable and γ is not. In the context of this statistical model segmentation is estimation of the set of labels.

Often one considers parametric models, in which the class-conditional probability functions $f_c(x)$ are assumed known, except for the values of distributional parameters. That is,

$$f_c(x) = f(x;\beta_c),$$

where β_c is the parameter. E.g., in the multivariate Gaussian case β_c consists of the mean and covariance matrix for class c.

In order to model the spatial correlation of images, one can assume that the labels γ_{ij} form a stochastic process, say a Markov process. One reads through the array in a fixed order, say, first row, left to right, second row, left to right, and conditions a given pixel on pixels preceding it in this ordering. Here a first-order Markov process would be one where a given pixel is conditioned on the pixels to the north and west of it. Thus the transition probability matrix has the following form for k=3 classes of segment.

Pixel to north	Pixel to west	Center pixel 1	2	3
1	1	P11,1	P11,2	P11,3
1	2	P12,1	P12,2	P12,3
1	3	P13,1	P13,2	P13,3
2	1	P21,1	P21,2	P21,3
2	2	P22,1	P22,2	P22,3
2	3	P23,1	P23,2	P23,3
3	1	P31,1	P31,2	P31,3
3	2	P32,1	P32,2	P32,3
3	3	P33,1	P33,2	P33,3

The total number of parameters, distributional parameters and transition probabilities, is large. But, with very large datasets one need not necessarily shy away from models with many parameters.

The algorithm developed for segmenting images according to this model is similar to that for segmenting time series, except now the transition-probability matrix is more complicated.

As a sample "image" the Fisher iris data were used. This dataset consists of 4 features measured on 150 flowers, 50 in each of three species. To form an image the 150 flowers were arranged into a 15 x 10 rectangular array, rows 1-5 being species 1, rows 6-10 being species 2, rows 11-15 being species 3. This means that the true segmentation is as follows.

TRUE SEGMENTATION:

ROW:	COLUMN:									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
6	2	2	2	2	2	2	2	2	2	2
7	2	2	2	2	2	2	2	2	2	2
8	2	2	2	2	2	2	2	2	2	2
9	2	2	2	2	2	2	2	2	2	2
10	2	2	2	2	2	2	2	2	2	2
11	3	3	3	3	3	3	3	3	3	3
12	3	3	3	3	3	3	3	3	3	3
13	3	3	3	3	3	3	3	3	3	3
14	3	3	3	3	3	3	3	3	3	3
15	3	3	3	3	3	3	3	3	3	3

Below are given results obtained by starting with initial means equal to the measurements on flowers 50, 100 and 150. (These are easy for the algorithm in the sense that they are in fact from the three different species, but not so easy as, e.g., flowers 1,51 and 101, which are further apart. Starting with means that are from correct classes is analogous to most applications, where something is known in advance about the characteristics of different classes of segment.) The results in successive iterations were as follows. Convergence was reached on the fourth iteration, i.e., on that iteration no pixel changed class.

SEGMENTATION ON ITERATION 1:

ROW:	COLUMN:									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
6	3	3	3	2	3	2	3	2	3	2
7	2	2	2	3	2	2	2	2	2	2
8	3	2	3	2	2	2	3	3	2	2
9	2	2	2	3	2	3	3	2	2	2
10	2	3	2	2	2	2	2	2	2	2
11	3	3	3	3	3	3	2	3	3	3
12	3	3	3	3	3	3	3	3	3	3
13	3	3	3	3	3	3	3	3	3	3
14	3	3	3	3	3	3	3	3	3	3
15	3	3	3	3	3	3	3	3	3	3

CONFUSION MATRIX:

		True Class			
		1	2	3	
Label	1	50	0	0	50
	2	0	36	1	37
	3	0	14	49	63
		50	50	50	150

SEGMENTATION ON ITERATION 2:

ROW:	COLUMN:									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
6	2	3	3	2	3	2	3	2	3	2
7	2	2	2	2	2	2	2	2	2	2
8	3	2	2	2	2	2	2	3	2	2
9	2	2	2	3	2	2	2	2	2	2
10	2	2	2	2	2	2	2	2	2	2
11	3	3	3	3	3	3	2	3	3	3
12	3	3	3	3	3	3	3	3	3	3
13	3	3	3	3	3	3	3	3	3	3
14	3	3	3	3	3	3	3	3	3	3
15	3	3	3	3	3	3	3	3	3	3

CONFUSION MATRIX:

		True Class			
		1	2	3	
Label	1	50	0	0	50
	2	0	42	1	43
	3	0	8	49	57
		50	50	50	150

SEGMENTATION ON ITERATION 3:

ROW:	COLUMN:									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
6	2	3	2	3	2	3	2	3	2	3
7	2	2	2	2	2	2	2	2	2	2
8	3	2	2	2	2	2	2	2	2	2
9	2	2	2	2	2	2	2	2	2	2
10	2	2	2	2	2	2	2	2	2	2
11	3	3	3	3	3	3	3	3	3	3
12	3	3	3	3	3	3	3	3	3	3
13	3	3	3	3	3	3	3	3	3	3
14	3	3	3	3	3	3	3	3	3	3
15	3	3	3	3	3	3	3	3	3	3

CONFUSION MATRIX:

		True Class			
		1	2	3	
	1	50	0	0	50
Label	2	0	44	0	44
	3	0	6	50	56
		50	50	50	150

Again, a good feature of the algorithms is that isolated labels tend to disappear, solving problems such as erroneously classifying an isolated pixel as corn in what is obviously a wheat field.

All computations reported here were carried out using FORTRAN computer programs written by the author. Tapes of these programs were sent to the Statistics Program at the Office of Naval Research (ONR) for deposit in the Computer Division of the Naval Research Laboratories. New versions of these programs, as well as other segmentation programs, are being developed. It is planned to publish these in Chen (1984).

Acknowledgement. This research was supported by Office of Naval Research Contract N00014-80-C-0408, Task NR042-443, and Army Research Office Contract DAAG29-82-K-0155, at the University of Illinois at Chicago.

References

Akaike, Hirotugu (1981). "Likelihood of a Model and Information Criteria." Journal of Econometrics 16, 1-14.

Ball, Geoffrey H., and Hall, David J. (1967). "A Clustering Technique for Summarizing Multivariate Data." Behavioral Science 12, 153-155.

Chen, C. H. (1984). Algorithms for Signal Segmentation, Change Detection and Spectral Estimation. Computer Science Press. Forthcoming.

Forney, G. David, Jr. (1973). "The Viterbi Algorithm," Proc. IEEE, vol. 61, pp. 268-278.

Nelson, Charles R. (1973). Applied Time Series Analysis for Managerial Forecasting. Holden-Day, Inc., San Francisco.

Parzen, E. (1982). "Maximum Entropy Interpretation of Autoregressive Spectral Densities." Statistics & Probability Letters 1, pp. 7 - 11.

Richards, J.A., Landgrebe, D.A., and Swain, P.H. "Overcoming Accuracy Deterioration in Pixel Relaxation Labeling," Proc. 5th International Conference on Pattern Recognition, Miam, Dec., 1980. IEEE Computer Society, Los Alamitos, CA, pp. 61-65.

Schwarz, Gideon (1978). "Estimating the Dimension of a Model." Ann. Statist., vol. 6, 461-464.

Sclove, Stanley L. (1983a). "On Segmentation of Time Series." In Studies in Econometrics, Time Series, and Multivariate Statistics: in Honor of T. W. Anderson (S. Karlin, T. Amemiya, and L. Goodman, eds.), Academic Press, pp. 311-330.

Sclove, Stanley L. (1983b). "Time-Series Segmentation: a Model and a Method. Information Sciences, vol. 29, 7-25

Southwell, R. (1940). Relaxation Methods in Bngineering Science: a Treatise on Approximate Computation. London: Oxford University Press.

Southwell, R. (1946). Relaxation Methods in Theoretical Physics. London and New York: Oxford University Press (Clarendon).

TABLE 1. GNP. Units: billions of current (non-constant) dollars (from Nelson (1973), pp. 100-101)

	Quarter							
	1	2	3	4	1	2	3	4
1947-48 GNP	224	228	232	242	248	256	263	264
change		4.0	4.2	10.3	5.9	7.6	6.9	1.4 -5.4
1949-50 GNP	259	255	257	255	266	27	293	305
change		-3.3	1.9	-2.1	11.0	9.4	17.7	11.4 13.5
1951-52 GNP	318	326	333	337	340	339	346	358
change		7.8	7.0	4.1	2.6	-0.4	6.5	12.1 6.5
1953-54 GNP	364	368	366	361	361	360	365	373
change		3.3	-1.7	-5.0	-0.1	-0.3	4.3	8.7 12.8
1955-56 GNP	386	394	403	409	411	416	421	430
change		8.2	8.1	6.3	1.8	5.6	4.4	8.9 7.4
1957-58 GNP	437	440	446	442	435	438	451	464
change		3.0	6.4	-4.8	-6.8	3.6	13.1	13.0 9.6
1959-60 GNP	474	487	484	491	503	505	504	503
change		12.9	-2.9	6.5	12.5	1.7	-0.5	-0..9 0.3
1961-62 GNP	504	515	524	538	548	557	564	572
change		11.3	9.3	13.5	10.1	9.4	7.2	7.6 5.4
1963-64 GNP	577	584	595	606	618	628	639	645
change		6.8	10.5	11.1	11.9	10.3	10.9	6.2 17.7
1965-66 GNP	663	676	691	710	730	743	756	771
change		12.9	15.4	18.9	19.5	13.8	12.6	14.8 13.5

TABLE 2. Fitting models

Model	AIC
Segmentation, 2 classes	481.4[a]
Segmentation, 3 classes, full trans. prob. matrix	483.6
Segmentation, 3 classes, sparse trans. prob. matrix[b]	488.5
Segmentation, 4 classes, full trans. prob. matrix	507.1
Segmentation, 4 classes, sparse trans. prob. matrix[b]	486.8
Segmentation, 5 classes, full trans. prob. matrix	506.5
Segmentation, 5 classes, sparse trans. prob. matrix	stopped[c]
Segmentation, 6 classes, full trans. prob. matrix	stopped[c]

a. Optimum, among segmentation models considered.

b. Allows transitions only to adjacent states.

c. Stopped, i.e., the algorithm reached an iteration where it allocated no observations to one of the classes.

TECHNICAL REPORTS

ARMY RESEARCH OFFICE CONTRACT DAAG29-82-K-0155

with the University of Illinois at Chicago

Statistical Models and Methods for
Cluster Analysis and Image Segmentation

Principal Investigator: Stanley L. Sclove

No. A82-1. Stanley L. Sclove. "Application of the Conditional Population-Mixture Model to Image Segmentation." 8/15/82

No. A82-2. Hamparsum Bozdogan and Stanley L. Sclove. "Multi-sample Cluster Analysis using Akaike's Information Criterion." 12/20/82

No. A82-3. Stanley L. Sclove. "Time-Series Segmentation: a Model and a Method." 12/22/82

No. A83-1. Hamparsum Bozdogan. "Determining the Number of Component Clusters in the Standard Multivariate Normal Mixture Model using Model-Selection Criteria." 6/16/83

No. A83-3. Stanley L. Sclove. "Use of Model-Segmentation Criteria in Clustering and Segmentation of Time Series and Digital Images." 5/5/83

6/23/83

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report No. UIC/DOM/A83-2	2. GOVT ACCESSION NO.: AD-A137 942 / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): On Segmentation of Digital Images using Spatial and Contextual Information via a Two-Dimensional Markov Model	5. TYPE OF REPORT & PERIOD COVERED: Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Stanley L. Sclove	8. CONTRACT OR GRANT NUMBER(s): ARO Contract Number DAAG29-82-K-0155
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Illinois at Chicago, Box 4348, Chicago, IL 60680	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709	12. REPORT DATE: December 6, 1983 13. NUMBER OF PAGES: 19 + ii
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

time-series segmentation, digital image segmentation, relaxation method, Viterbi algorithm, model-selection criteria, information criteria, Akaike's information criterion, Schwarz' criterion

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(Please refer to next sheet.)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The problem of partitioning a digital image into segments is considered. First the procedure is illustrated for the analogous one-dimensional problem, namely, segmentation of time series. Then similar ideas are applied to the segmentation of digital images.

The segments are considered as falling into classes. A probability distribution is associated with each class of segment. Parametric families of distributions are considered, a set of parameter values being associated with each class. With each observation is associated an unobservable label, indicating from which class the observation arose. The label process is modeled as a Markov process. Segmentation algorithms are obtained by applying a method of iterated maximum likelihood (a relaxation method) to the resulting likelihood function. In this paper special attention is given to the situation in which the observations are conditionally independent, given the labels. Numerical examples are given. Choice of the number of classes, using statistical model-selection criteria, is illustrated.

END

FILMED

3-84

DTIC